Zijlstra-Shaw, S., Roberts, T., & G Robinson, P. (2017). Evaluation of an assessment system for professionalism amongst dental students. *European Journal of Dental Education*, *21*(4), e89–e100. https://doi.org/10.1111/eje.12226

Evaluation of an Assessment System for Professionalism amongst Dental Students

Introduction

Dental professionalism is an essential requirement to practice dentistry in the current environment[1]. It is a construct that covers both abilities and personal qualities[2]. There is therefore a need for programmes of assessment within dental education that are both formative, to aid professional development, and summative, to determine whether students have reached appropriate standards.

Methods and systems that have been used to assess medical professionalism are well documented[3-8] and may be useful within dentistry. Unfortunately no single method has been found that adequately measures all aspects of professionalism[8].

Assessment must be based on explicit criteria[9-10] and on an agreed definition and framework alongside a suitable model of learning[2,11]. . Furthermore, the tools used for assessment must be evaluated to ensure they are fit for the purpose.

Our previous review and qualitative study arrived at a definition of professionalism as "*the manner in which one reflects on and reconciles different aspects of practice which demonstrates acceptance of responsibility and accountability. It is manifested in the manner in which work is carried out"* [12]. However definitions alone are inadequate to assess professionalism. They are criticized for being both too much and too little[13] and being lists of idealistic values that may be both inadequate and vague[14].

Frameworks are useful in education to drive learning, provide clarity and guide observation[15]. This is particularly important when assessing professionalism, as observation alone has been found to be inadequate[16-17]. Our definition was operationalised into a conceptual framework comprising 8 domains understands self, understands others, trustworthiness, ability to relate to context, vocational, altruistic, reliability and accountability all harmonised through reflection; to shape an assessment system[12]. The framework was

then used alongside a model of assessment in learning[18] to produce aims and objectives within an existing curriculum. During panel testing the framework was modified to ensure it was relevant to dental students (Figure 1) and then used to produce an assessment system, to be used by staff and senior dental students within one area of the curriculum. This Assessment of Dental Student's Professionalism System (ADSPS) was then piloted amongst a cohort of dental students.

Evaluation of this system was required to determine the validity and reliability of the system and the validity of the underlying framework. Therefore this study aimed to evaluate the system to assess dental students' professionalism.

Method

The evaluation consisted of three aspects; qualitative panel testing before piloting, qualitative evaluation of ADSPS during piloting and quantitative evaluation by analysis of students' marks during piloting.

Panel testing

Before piloting the ADSPS was evaluated by panel testing participants' opinions of its feasibility, acceptability and face and content validity. Panel members included senior staff, (dentally and non-dentally qualified) along with student representatives of a dental school. Three focus groups were arranged, each with 4-8 participants. The student focus group was held separately from the staff, so that both groups could speak openly. The data were recorded, transcribed and analysed using content analysis.[19] Suggestions for modifying the ADSPS were agreed by consensus.

Piloting

Within the existing curriculum senior dental students attend three primary care "Outreach" placements in either National Health Service general practices (6) or salaried dental services

(5). Each placement receives between 2 and 5 students for six weeks throughout the second half of the students' fourth and first half of their fifth year of study.

During the placement students provide dental care commensurate with the nature of the placement. The general practices provide comprehensive dental care for patients, many of whom are long term patients of the practice, within the regulations laid down by the NHS for general practice[20]. The salaried dental services treat children and patients needing special care and included an access centre caring for patients experiencing difficulty accessing routine care.

All placements are small establishments and students work under close supervision, provided continuously by both dental nurses (within their scope of practice) and by dentally qualified clinicians. The ratio supervising dentists to students varies from 1:2 to 1:5.

The students attend three placements so that their experience accumulates over a wide range of both patients and procedures[20], allowing them to learn and demonstrate their abilities in different contexts. Outreach was therefore considered a suitable part of the curriculum to pilot the ADSPS.

Placement staff were approached for this evaluation initially by e-mail, followed by placement visits, during which training and calibration was provided.

The students were given details of the pilot during a specifically designed lecture, at which they were encouraged to ask questions about the system. All students and practices were also given a handbook explaining the ADSPS.

At the end of each placement students reflected on their experiences and both gave and received specific feedback. Staff collected feedback and encouraged reflection on students' personal learning. The ADSPS consisted of an appraisal meeting between each student and a supervising clinician at each placement. Students were assessed using three forms based on observations during the 6 weeks of the placement. (Figure 2).

At the end of each placement students assessed their own performance on a form containing 8 items relating to professionalism. The supervising clinicians completed an analogous form based on a combined view from all members of staff at the placement. These documents then formed the basis of a discussion with individual students about their performance. The assessments were then recorded on a combined results form (agreed form) comprising 16 items (see appendix 3). A global rating of professionalism was also collected separately for each student. During the evaluation the results were used to provide formative feedback only.

The forms were designed to record observed behaviour over time, whilst the appraisal meeting allowed modification, where students could give an account of their behaviour. Thus, observed behaviours and the second order nature of professionalism were taken into account and the students were encouraged to develop their skills in both overt and tacit aspects of professionalism by reflecting on staff feedback.

Each item was scored on a six point Likert scale where scores of 1 and 2 denoted performance below the level expected from students at their stage of training. The middle scores (3, 4) denoted performance at the level expected, and the upper range (5,6) denoted that well above the level expected. Descriptors were provided to guide staff and students when completing the forms.

Approval for the pilot was obtained from the University of Sheffield Research Ethics Committee before participants were approached or recruited.

Quantitative evaluation

The response process, which describes the way an assessment system is used by the assessors and assessed, was evaluated by descriptive analysis and by comparing the scores by students and staff over eight items.

Internal reliability was evaluated via Cronbach's alpha, alphas with each item deleted and corrected item-totals, on the agreed assessment form from the first available complete data set.

Test-retest reliability was evaluated by comparing the first and second assessments for each student for which there were complete data. These scores were related to the student and not the tutor, as the student placements were allocated randomly by the School administration. Intraclass correlation coefficient (ICC) was calculated for each item using a one-way repeated measures model.

Both construct and criterion validity were evaluated. Construct validity was evaluated by assessing correlations between items within the model hypothesized to be related or unrelated (see figure 1) based on the following hypotheses;

- There would be a correlation between "self-awareness" and "reflection"
- There would be a correlation between "self-awareness" and "vocational" aspects, especially self-motivation
- There would be a correlation between "awareness of others" and "altruism" especially caring and respect
- There would be a correlation between "trustworthiness" and "responsibility"
- There would be a correlation between "ability to relate to context" and "accountability"

and the following hypothesis;

- There would be no associations between the scores and age or gender.

Correlations were evaluated using Spearman rank correlation coefficients from the agreed assessment form.

Criterion validity was calculated by correlating (Spearman's rho) each item with the global rating using data from the agreed assessment form.

All statistical analysis was carried out using SPSS 20.

Qualitative evaluation of the system

The feasibility and acceptability of ADSPS were evaluated qualitatively by the students at their regular feedback sessions. During these sessions the students divide into a group for each placement to provide written and verbal feedback. The students were encouraged to provide verbal feedback on the assessment system and notes were taken. Students wishing to provide more detailed feedback were asked to e-mail the researcher. The data were analysed using content analysis[19].

Results

Panel testing

Themes arising from the panel tests included improved assessment criteria, greater ability to reward positive student behaviour and the need for staff training.

The framework was described as useful as it helped clarify educational aims and objectives. It also helped increase awareness of the importance of professionalism, and was thought to encourage reflective practice. Student learning and staff expectations were also seen to be standardised by this clarification, which allowed positive student behaviour to be rewarded and was anticipated to increase the effects of students' socialisation.

The resource implications and effects on the institutional culture were seen as a challenge, but one which could be met by appropriate staff calibration and training so that the staff could more easily understand the processes concerned. Staff training was also proposed to increase reliability within the system. The only negative comments referred to the resource implications of implementing the system.

All the domains described in the model were agreed to be relevant. No omissions were noted. Feasibility, acceptability and face and content validity were endorsed, although some

modification of the language to make it more consistent with that more commonly used by staff and students was agreed by consensus.

The language was modified and cross-checked with both the framework and the data underlying the framework to ensure this remained aligned

Quantitative evaluation

The cohort consisted of 81 students of whom 44 were female. Ages ranged from 21 to 38 years with a median of 23 years. Each student attended 3 placements; however, problems were encountered at the inception of the pilot with some placements missing the start of the study. Nevertheless data were collected for 81 students at either their first or second placement, referred to as the first complete data set. Follow up data were collected for 66 students at a second placement. The forms were completed by 19 different members of staff in 9 different placements.

Descriptive analysis of the first complete set of data showed that students used a slightly narrower range of scores than staff, with students scoring between points 3 and 6, whilst staff used points 2-6 inclusive. The modal rating was a score of 4 used in 56.9% of the cases by students, 55.7% by staff and 58.4% of the combined forms (Figure 3).

Staff, student and agreed scores were correlated for all but two domains (Table 1). Only staff vs student scores for *consideration* and *relating to context* were not significantly correlated. Thus, the agreed form could be seen to reflect the joint views of staff members and students. This agreed form was then used for the further analysis.

Internal consistency was evaluated using item total correlations and Cronbach's alpha coefficient in the first complete set of data. All item-total correlations exceeded 0.65 and Cronbach's alpha, based on all 16 standardised items, was 0.95, which was taken to mean

that internal consistency was very high. This consistency persisted for each item or domain when deleted (Table 2),

The ICC was calculated using the data from 66 students for whom follow up data were available. ICCs above 0.9 indicate that the measure is stable over time. "*Reflection shows balance*" was the only variable with scores under this threshold, with a value of 0.88. The remainder had values between 0.96 and 1 (Table 3).

Construct validity was evaluated by assessing correlations between items within the framework hypothesized to be related or unrelated. All the hypothesised correlations were significant, all but one being r_s >0.5, p <0.01 (Table 4).

Correlations between the scores and age or gender were hypothesized to be unrelated and there was no significant difference for gender where P=0.28, Mann Whitney U test, (Table 5), Students' age was unrelated to "*commitment*", "*consideration*", "*responsibility*", "*understands abilities*", "*instils trust*", "*accounts transparently*" and "*respects rules*", (r_s = 0.02-0.19, p 0.1 -0.9, Table 6). Age was weakly related to "*trustworthiness*", "*relates to context*", and the three domains relating to the appropriate manifestation of student reflection (r_s0.22 - 0.32 p <0.05, Table 6).

Criterion validity was confirmed as each domain on the agreed form correlated with the global ratings provided by the supervising clinicians (Table 7, all r_s>0.32 and p< 0.05).

Qualitative evaluation

Data from the meetings held after ADSPS had been used in each Outreach placement, were recorded with contemporaeous notes during the meetings. Nine students (11%) supplemented those data with further e-mails.

Content analysis produced three main themes; the ADSPS process, educational value and suggestions for improvements. The results are summarised in these themes using pseudonymised quotes for illustration.

The ADSPS process was initially found to be confusing. Students commented that the forms were "*quite complicated and had to refer to the descriptors a lot*" (Helen). However with time some students reported that they "*understood them better*" (Mary) and found them "*interesting to do*" (George). By the second or third placement, it was reported that staff were better able to complete the forms, stating that "*as they have more practice and this is making the feedback from the forms more useful*" (Sean).

Training was seen as a way of improving understanding of the ADSPS as

> *"The staff at the practice were very complimentary of the Outreach Training day(s) ... the format and content of the day was constructive and it was the most interesting and enjoyable of the day"* (Sarah).

This necessity of having staff members committed to the system was also commented on by another student;

> *"Essentially, it mustn't be forgotten that for the system to work, it is heavily reliant on the tutors/supervisors' professionalism and their desire to go the extra mile, providing extensive constructive criticism in an articulate, non-judgmental fashion."* (Philipa)

The importance of allowing sufficient time to complete the system was reported alongside inferences that when it was rushed and not completed properly it did not appear to be worthwhile;

> "*it was important to put time aside to complete the forms and they could not be done in a rush*" (Chris).

However, reports from placements where time was put aside and the forms were completed properly suggested that it was a valuable exercise, commenting on one tutor that she *"was very good and blocked off time to complete the forms."* (Sean).

Furthermore "*the one-on- one feedback session conducted at the end of the placement was an ideal way to communicate the results of the assessment".* (Adnan).

The third aspect of the system some students particularly like was the ability of the ADSPS to involve all staff members. One student commented that

> "*This meant that our professionalism was assessed on the basis of our performance throughout the entire placement, this would not be possible if only one member of staff was responsible for the assessment*" (Moira)

This flexibility to involve all staff and then combine the results was seen to offer advantages as it enabled the placement to form a more complete picture of the students' professionalism. This aspect of the system meant one supervising clinician *"had gained an accurate and complete idea of our professional performance during the placement"* (Blake).

Thus, the system was seen to be initially complex, requiring both staff and student training, and whilst time consuming, it was found to be useful.

The educational value of the system was seen in the increased recognition of the importance of learning professionalism and the formalisation of the feedback provided by staff.

"The system provides better more constructive feedback" (Jane)

and

"It was a great idea to do this, because the feedback I received was in depth, comprehensive and gave a great insight to my strengths as well as areas in which I could improve on". (Philip).

Some students found it useful to compare their self-ratings with those of the staff. Explanations on how to improve "*allowed his tutor to pin point an area he wasn't aware was a problem and give him advice of how to improve*" (Simon).

This ability of the ADSPS to allow feedback prompted student reflection. One student commented that ADSPS "*helped me evaluate how other professionals perceive me and helped me identify particular areas I needed to work on*" (Clare).

ADSPS was also felt to be useful and appropriate on Outreach as

> *"it was a very good exercise because in general practice, professionalism is much more central to one's patient management and team working"* (Jenny)
>
> and
>
> *"everyone I've spoken to feels it is a positive contribution to the assessment of our time on outreach"* (Alan).

This combination of feedback and reflection was seen as important for students' professional development. One student had saved the negative comments received at the end of his first placement onto his mobile phone. He then looked at the comments every day during his second placement and worked to improve on weaker areas. At the end of his second placement he got much more positive comments and had improved.

After the pilot both staff and students were encouraged to suggest improvements as part of the evaluation. Suggestions included improvements to the categories. One student had particular problems with one area of the form commenting;

> *"Commitment, autonomy, confidence and motivation are in the same row. Taking myself as an example my tutors found I lacked some confidence and marked me lower in this row. Therefore this also lowered my motivation and commitment score even though I had shown no lack of these"* (Anisha)

However, this was tempered by understanding the practicalities of the situation

"I understand that this is hard to fit onto the one sheet" (Anisha).

Despite suggested improvements to the clarity of the wording on the forms, students commented that "*We felt the guide covered all aspects of professionalism, nothing was left out."*

Discussion

The aim of this study was to evaluate ADSPS when piloted amongst senior dental students attending Outreach placements. The qualitative data suggest that the ADSPS provides useful feedback and encourages student reflection. The quantitative data reveals ADSPS to have good psychometric properties. Thus this preliminary study suggests that the ADSPS is a feasible system which is both reliable and valid and has educational value when used to assess dental students' professionalism.

Face and content validity of ADSPS were evaluated by panel testing before the assessment system was used, consequently some modification of the language was recommended and amended. However, notwithstanding these recommendations, initial comments suggested improvements could still be made to the wording on the forms.

Analysis of the awarded marks demonstrated good alignment between student and staff interpretations with the agreed form reflecting the joint views of staff members and students. The marks awarded ranged from point 2-6, demonstrating the use of most points on the scale and thus the potential for discrimination between students and over time. Scores were positively skewed towards the upper end of the scales (Figure 3). This skewness has educational value as it may help identify borderline and failing students, who are often not recorded by other systems[21] and may reflect the developed professionalism of these senior

students. Furthermore, professionalism is often assessed using a global score and staff aggregate good points with less good ones, thus borderline behaviour isn't often recorded.

The piloted system divided professionalism into its constituent parts and that, in combination with the 1-6 point scale, enabled the less able students to be identified. Indeed 8.6% of the scores given by staff were a '3'. Moreover, by breaking professionalism into its constituent parts, staff may have become more aware of the different domains involved and thus been encouraged to recognise lower professionalism in specific areas.

The reliability of ADSPS was consistently high. The item-total correlations (all > 0.6) and Cronbach's alpha (>0.95) (Table 2) exceed standards for assessment instruments for individuals suggesting that the system could be useful for summative assessment[22]. Those high values, even with each item removed indicate that all the items are construct relevant. Whilst many domains are assessed by ADSPS, any reduction could reduce its content validity. In addition, it would reduce the educational value by reducing the specificity of the feedback, an aspect of ADSPS that the students valued. Thus, despite the high degree of reliability, using the ADSPS summatively could have a negative effect on the value of the feedback and the openness of the students' reflection, which could compromise the educational value of the system.

Test-retest reliability was very high (Table 3), supporting the reliability of ADSPS as the forms were completed by a total of 19 different members of staff across 9 different Outreach placements. Due to the high number of placements and tutors, Generalizability theory was considered to calculate reliability, however the breadth of data was not available to ensure meaningful interpretation of the results and ICC is considered a legitimate test under these circumstances[22].

Construct validity of ADSPS was confirmed in statistical relationships between related domains within the framework. The slightly lower values related to context may reflect some students adapting to their new environments better than others.

The convergent validity hypotheses that there should be no associations between professionalism scores and age or gender were supported (Tables 5 and 6).

Each domain on the agreed form correlated with the global ratings provided by the supervising clinicians (Table 7) with all but two items being r_s >0.4. Of the domains with a lower value of r_s, qualitative data show that "*trustworthiness*" was interpreted differently by different placements.

The qualitative data affirm the educational impact and acceptability of the ADSPS, with it said to make a positive contribution to student assessment on Outreach.

Time factors were also seen to be important. The quality of the feedback given was time dependent. This has cost implications, particularly in situations where staff time is expensive, which may affect the feasibility of implementing ADSPS broadly across a curriculum. However, the importance of protecting time to teach, despite heavy clinical workloads, is stressed within the literature relating to role modelling and the teaching of professionalism in medical education[23-24]. Thus, this needs to be recognised as an important aspect within the curriculum for undergraduates.

The appraisal meetings allowed discussions that enabled exploration of the reasoning behind some of the students' actions. Students described this as very useful for encouraging reflection and providing appropriate feedback. Furthermore this accords with recommendations to assess professionalism using both observed behaviour and the reasoning behind this[17].

Some students found it useful to compare their self-ratings with those awarded by staff and to receive explanations on how to improve. This prompting is also seen in medical education[25]. In addition, including many staff members within one assessment was seen to create a broader picture of the student's professionalism, which also accords with the literature on 360º feedback[26].

The specificity of the feedback and its ability to stimulate reflection were seen as important as part of students' professional development. This again accords with the literature on reflection within medical education[27-29], which describes the strong link between good feedback and reflection. Furthermore the encouragement of reflective practice is seen as an important aspect of professionalism[30]. ADSPS encourages reflection by making this aspect explicit and making the link between reflection and other competences such as communication and clinical skills by recognising the second order nature of professionalism. This is a key component of our definition of professionalism.

However, individual aspects of assessment, feedback and reflection tend to have been reported separately and reports of the psychometric properties of any measures are limited[5,7]. A lack of theoretical models has limited any medical curriculum design with respect to this integration and assessment of professionalism[30]. ADSPS not only accords with these individual aspects, it integrates them within a framework that allows them to be identified in a structured manner.

This pilot indicated the need for staff training and calibration, for evaluation and to involve the assessors in the wording of the descriptors to ensure that they are aligned to the reality map of the assessors[31]. Further refinement may be needed to ensure the clarity of the descriptors and their content validity for the varying stages of student education from beginner to competent.

Although the numbers of students assessed was small, the analysis showed the ADSPS to be reliable and valid and also demonstrated the strength of the underlying conceptual framework. Furthermore the qualitative evaluation demonstrated the usefulness of ADSPS alongside its acceptability, feasibility and educational impact. Thus, ADSPS is theoretically and empirically robust. ADSPS also allowed students to receive detailed feedback and encouraged student reflection.

This pilot study involved one cohort of students from one dental school. Further research is needed to confirm the generalisability of ADSPS to other curricula. This will require the assessment to be adapted to any new context. However the current study involved a broad range of primary care placements, thus the system may be generalisable.

The number of staff involved, even in this pilot study brought advantages; in that it was inclusive and allowed views from nursing staff as well as supervising clinicians. It also demonstrated staff training and calibration to be feasible.

The purpose of the pilot was to evaluate an assessment system based on previously developed framework. All aspects of the framework were incorporated into the ADSPS. This development was ambitious. Whilst there was general understanding that professionalism was complex and that covering it comprehensively was commensurately complex, the evaluation suggests that covering different aspects in different contexts could also be appropriate.

Conclusion

This evaluation reveals ADSPS to have good internal reliability and validity and suggests that basing an assessment system around the model developed in phase one of this research is a valuable approach to the assessment of professionalism within dental education.

References

1. Cowpe J, Plasschaert A, Harzer W, Vinkka-Puhakka H, Walmsley AD. Profile and Competences for the European Dentist - update 2009 Association for Dental Education in Europe; Available at http://www.adee.org/cms/uploads/adee/TF_I_V2_September2009.pdf .[Accessed 28th February 2010].

2. Zijlstra-Shaw S, Roberts TE and Robinson PG. (2012) Assessing Professionalism within Dental Education; the need for a definition *European Journal of Dental Education* 16(1):e128-36

3. Arnold, L. (2002) Assessing professional behaviour: Yesterday, Today, and Tomorrow *Academic Medicine* 77, 502-515

4. Lynch, D.C., Surdyk, P.M. and Eiser, A.R. (2004) Assessing professionalism: a review of the literature *Medical Teacher* 26(4), 366–373.

5. Veloski J. J., Fields S.K., James R.N., Boex R. and Blank L.L., (2005)Measuring Professionalism: A Review of Studies with Instruments Reported in the Literature between 1982 and 2002 *Academic Medicine* 80,366–370

6. Stern, D. T. (ed.) (2006) *Measuring Medical Professionalism* Oxford, Oxford University Press

7. Jha, V., Bekker, H. L, Duffy, S.R.G. and Roberts,T. E. (2007) A systematic review of studies assessing and facilitating attitudes towards professionalism in medicine *Medical Education* 41,822-829

8. Van Mook, W.N.K.A., Gorter, S.L., O'Sullivan, H., Wass, V., Zwaveling, J.H., Schuwirth, L.W.T. and van der Vleuten, C.P.M. (2009) Approaches to professional behaviour assessment : Tools in the professionalism toolkit European Journal of Internal Medicine20, e153–e157

9. Stern, D.T. and Papadakis, M. (2006) The Developing Physician - Becoming a Professional *The New England Journal of Medicine* 355(17), 1794-1799

10. Hawkins, R.E., Katsufrakis, P.J., Holtman, M.C. and Clauser, B.E. (2009) Assessment of medical professionalism: Who, what, when, where, how, and . . .why? *Medical Teacher* 31, 348–361

11. O'Sullivan, H., van Mook, W., Fewtrell, R. and Wass, V. (2012) Integrating professionalism into the curriculum: AMEE Guide No. 61 *Medical Teacher* 34, e64–e77

12. Zijlstra-Shaw S, Roberts TE and Robinson PG. (2013) Perceptions of Professionalism in Dentistry - a Qualitative Study *British Dental Journal* 215 E18 1-6

13. Erde, E. L. (2008) Professionalism's Facets: Ambiguity, Ambivalence, and Nostalgia *Journal of Medicine and Philosophy* 33, 6-26

14. Wynia MK, Papadakis MA, Sullivan WM and Hafferty FW (2014) More Than a List of Values and Desired Behaviours: A Foundational Understanding of Medical professionalism Academic Medicine 89;712-714

15. Pangaro, L. and Ten Cate, O. (2013) Framework for learner assessment in medicine: AMEE Guide 78 *Medical Teacher* 35, e1197-e1210

16. Ginsburg, S, Regehr G., and Lingard, L. (2004) Basing the evaluation of professionalism on observable behaviors: a cautionary tale *Academic Medicine* 79, S1–4.

17. Ginsburg S, Regehr, G. and Mylopoulos, M. (2009) From behaviours to attributions: further concerns regarding the evaluation of professionalism *Medical Education* 43, 414–425

18. Miller, G.E. (1990) The Assessment of Clinical Skills/Competence/Performance *Academic Medicine* 65(9 Suppl),S63–S67.

19. Bryman, A. (2008) *Social Research Methods* chapter 12, p273-293 Oxford, Oxford University Press

20. Smith M, Lennon M. A. and Robinson P. G (2010) Students' clinical experience on outreach placements *Eur J Dent Educ* 14 7–11

21. Van Mook, W.N.K.A., De Grave, W.S., Huijssen-Huisman, E., De Witt-Luth, M., Dolmans, D.H.J.M., Muijtjens, A.M.M., Schuwirth, L.W.T. & Van der Vleuten, C.P.M, (2007)Factors inhibiting assessment of students' professional behaviour in the tutorial group during problem-based learning *Medical Education* 41, 849–856.

22. Downing SM (2004) Reliability: on the reproducibility of assessment data *Medical Education* 38: 1006–1012

23. Wright, S.M. Kern, D.E. Kolodner, K. Howard, D.M. and Brancati, F.L. (1998) Attributes of Excellent Attending-Physician Role Models *New England Journal of Medicine* 339,1986-93

24. Cruess, S.R. Cruess, R.L. and Steinert, Y. (2008) Role modelling – making the most of a powerful teaching strategy *British Medical Journal* 336.718-721

25. Rees, C. and Shepherd, M. (2005) Students' and assessors' attitudes towards students' self-assessment of their personal and professional behaviours *Medical Education* 39, 30–39.

26. Rees, C. and Shepherd, M. (2005) The acceptability of 360-degree judgements as a method of assessing undergraduate medical students' personal and professional behaviours *Medical Education* 39, 49–57

27. Atkins, S. and Murphy, K. (1993) Reflection: a review of the literature *Journal of Advanced Nursing,* 18(8), 1188-1192

28. Boenink, A.D. (2006) *Teaching and learning reflection on medical professionalism* PhD Free University of Amsterdam

29. Muir F, Scott M, McConville K, Watson K, Behbehani K and Sukkar F (2014) Taking the learning beyond the individual: how reflection informs change in practice *International Journal Of Medical Education* 5: 24-30

30. Passi, V. Doug,M. Peile, E. Thistlethwaite,J. and Johnson,N. (2010) Developing medical professionalism in future doctors: a systematic review *International Journal of Medical Education* 1.19-29

31. Crossley, J. and Jolly, B. (2012) Making sense of work-based assessment: ask the right questions, in the right way, about the right things of the right people Assessing health professionals" *Medical Education* 46, 28–37.